

(43) Date of publication of application: 19.08.97

G06F 17/30
G06F 12/00

(71) Applicant: **NEC CORP**

(72) Inventor: **KUBO SHINYA**

the origin are searched in a required range.

COPYRIGHT: (C)1997,JPO

SOLUTION: The number of the nodes from the node to be an origin to the node of a search destination is turned to a hierarchy number and the limiting conditions of a search range are set by the maximum value of the hierarchy number. A file collecting part 21 traces a link from the node to be the origin, successively reads the nodes from a server machine 13, obtains the title of the link destination node, a storage position and the correspondence relation of a link origin and a link destination from the nodes and registers them. A hierarchy calculation part 27 obtains the hierarchy number of the link destination node of the node every time the node is read from the server machine 13. When the hierarchy number becomes more than the set maximum value, the tracing of the further link and the read of the nodes are stopped. Since the search range is limited by the hierarchy number from the origin, only the nodes with the semantic connection with the node of

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平9-218876

(43)公開日 平成9年(1997)8月19日

(51)Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/30			G 0 6 F 15/40	3 1 0 C
12/00	5 4 5		12/00	5 4 5 A

審査請求 有 請求項の数4 O L (全 16 頁)

(21)出願番号 特願平8-22343

(22)出願日 平成8年(1996)2月8日

(71)出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72)発明者 久保 信也

東京都港区芝五丁目7番1号 日本電気株式会社内

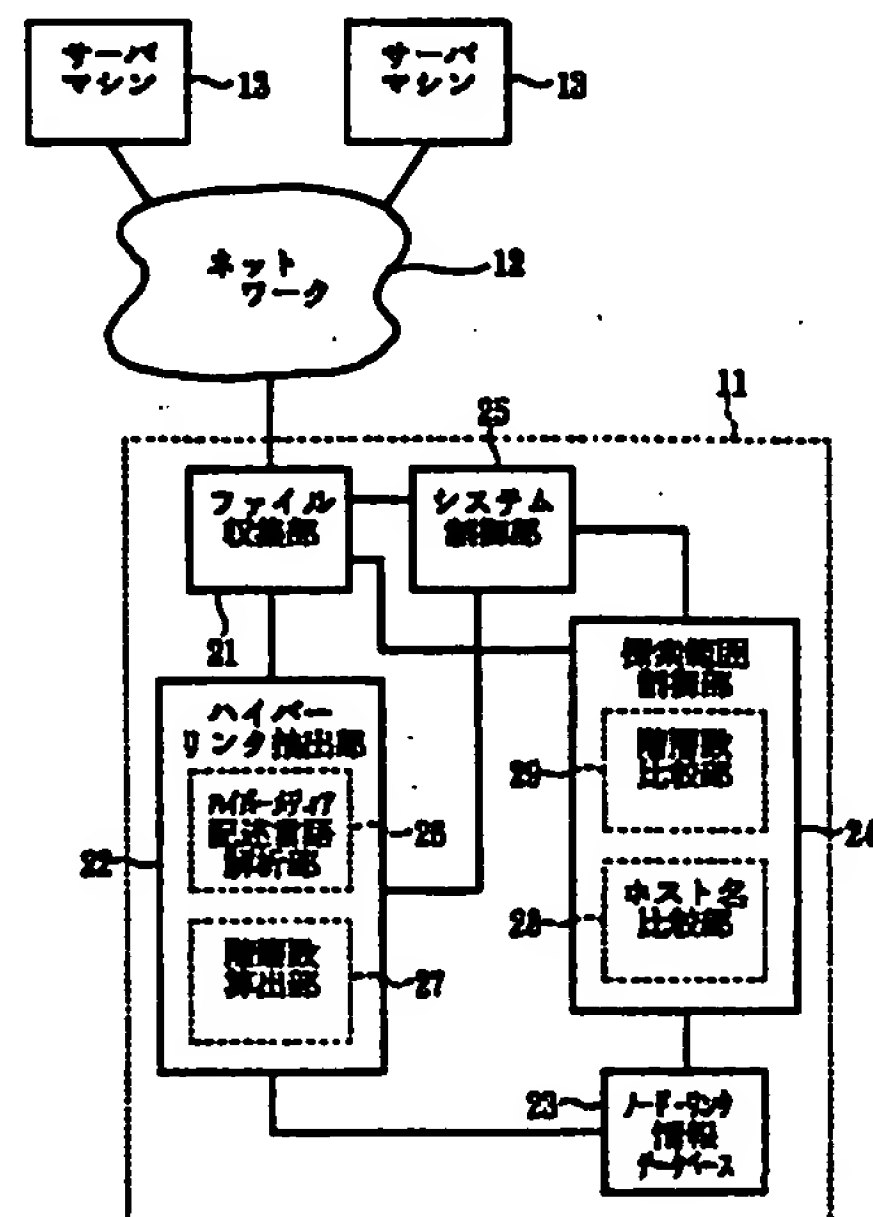
(74)代理人 弁理士 山内 梅雄

(54)【発明の名称】 ノード・リンク探索装置

(57)【要約】

【課題】 意味的なつながりのない不要なノードの探索を制限することのできるノード・リンク探索装置を提供する。

【解決手段】 起点となるノードから探索先のノードまでの間のノードの数を階層数とし、探索範囲の制限条件を階層数の最大値により設定する。ファイル収集部21は起点となるノードからリンクをたどりサーバマシン13より順次ノードを読み出し、このノードからそのリンク先ノードの名称や格納位置やリンク元とリンク先の対応関係を取得して登録する。階層算出部27はノードをサーバマシン13から読み込むごとにそのノードのリンク先ノードの階層数を求める。階層数が設定した最大値以上になったとき、それ以上先のリンクをたどりノードを読み出すことを中止する。起点からの階層数で探索範囲が制限されるので、起点のノードと意味的なつながりのあるノードのみを必要な範囲で探索できる。



【特許請求の範囲】

【請求項1】 ハイパーテキストの各ノードに含まれるリンク先のノードの名称とリンク先のノードの格納位置とを表わしたリンク情報を基にして任意のノードからリンク先のノードへの探索を順次行う際の探索範囲の制限条件を探索の起点となるノードから探索先のノードまでの間に存在するノードの数である階層数の最大値として設定する探索条件設定手段と、

前記リンク情報の示すリンク先のノードの内容をそれを蓄積しているサーバから読み出すことを前記探索の起点となるノードから順に繰り返し行うファイル収集手段と、

このファイル収集手段によって1つのノードの内容を読み込むごとにそのノードの含むリンク情報およびこの示すリンク先のノードと今回読み込んだノードとの対応付けを表わす情報とを記憶するノード・リンク情報記憶手段と、

前記ファイル収集手段によって1つのノードの内容を読み込むごとにそのノードに含まれるリンク情報の示すリンク先のノードの階層数を求める階層数算出手段と、

この階層数算出手段によって求めた階層数が前記探索条件設定手段により設定した階層数の最大値よりも大きいとき今回読み込んだノード以降にリンクされているノードの内容の前記ファイル収集手段による読み込みを中止させる探索範囲制限手段とを具備することを特徴とするノード・リンク探索装置。

【請求項2】 ハイパーテキストの各ノードに含まれるリンク先のノードの名称とリンク先のノードの格納位置とを表わしたリンク情報を基にして任意のノードからリンク先のノードへの探索を順次行う際の探索範囲の制限条件を探索の起点となるノードから探索先のノードまでの間に存在するリンクの数である階層数の最大値として設定する探索条件設定手段と、

前記リンク情報の示すリンク先のノードの内容をそれを蓄積しているサーバから読み出すことを前記探索の起点となるノードから順に繰り返し行うファイル収集手段と、

このファイル収集手段によって1つのノードの内容を読み込むごとにそのノードの含むリンク情報およびこの示すリンク先のノードと今回読み込んだノードとの対応付けを表わす情報とを記憶するノード・リンク情報記憶手段と、

前記ファイル収集手段によって1つのノードの内容を読み込むごとにそのノードに含まれるリンク情報の表わすリンク先のノードへのリンクの階層数を求める階層数算出手段と、

この階層数算出手段によって求めた階層数が前記探索条件設定手段により設定した階層数の最大値よりも大きいとき今回読み込んだノード以降にリンクされているノードの内容の前記ファイル収集手段による読み込みを中止

させる探索範囲制限手段とを具備することを特徴とするノード・リンク探索装置。

【請求項3】 探索の対象となるノードがネットワークに接続された複数のサーバに分散して格納されていることを特徴とする請求項1または請求項2記載のノード・リンク探索装置。

【請求項4】 前記ファイル収集手段は、今回読み込んだノードに含まれているリンク情報の示すリンク先のノードが既に読み込んだノードと同一であるか否かを判別する同一ノード判別手段と、この同一ノード判別手段により既に読み込んだノードと同一であると判別されたときそのノードの再度の読み込みを中止する多重読み込み中止手段とを具備することを特徴とする請求項1または請求項2記載のノード・リンク探索装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ノードとリンクからなるデータベースにおいて指定された条件の下で探索の対象とすべきノードの範囲を見出すノード・リンク探索装置に係わり、特に分散ハイパーメディアシステムのようにノードがネットワーク上の複数のサーバに分散して格納されたデータベースで用いられるノード・リンク探索装置に関する。

【0002】

【従来の技術】データベースの1つとして、ドキュメントファイルなど各種情報を格納した複数のノードをこれらの従属関係を表わしたリンクをたどることによって検索を行うハイパーテキストシステムがある。このようなハイパーテキストでの検索作業は、ノードに格納されている情報のインデックスと、ノード相互間のリンク関係を予め保持しておくことでスムーズに進めることができる。

【0003】特開平4-321144号公報には、ノード間相互間の関係を容易に把握できるように表示できるハイパーテキストシステムが開示されている。各ノードは任意のノードにリンクすることができるので、複数のノード間でループ状にリンクが形成されることがある。このシステムでは、ノード相互間のリンク関係を予め登録したテーブルを基にして表示画面を作成している。そして、ループ状になっているリンクを切り離し、あるノードを起点としたときに木構造としてノード相互間のリンク関係が表示されるようにしている。

【0004】従来、ハイパーテキストシステムはローカルのワークステーションなどに構築されていたが、近年の通信技術の発達により、ネットワークを介して接続された複数のサーバにノードを分散して格納するものが登場している。このようなシステムは、分散ハイパーメディアシステムと呼ばれている。たとえば、インターネット上の情報発信手段として、ワールド・ワイド・ウェブ (World Wide Web 以下WWWと表わす。) が注目され

ている。ハイパーメディアとは、文字や表などのテキストデータだけでなく、動画や音声などのマルチメディアデータも扱うことのできるハイパーテキストのことである。

【0005】分散ハイパーメディアシステムでは、各サーバに格納されているノードの情報をネットワークを介して取得することによって、ノード相互間のリンク関係を表わしたテーブルを作成するようになっている。ノード相互間のリンク関係を表わした情報をノード・リンク情報データベースと呼ぶことにする。また、ノード・リンク情報データベースを作成する装置をノード・リンク探索装置と呼ぶことにする。

【0006】ワールド・ワイド・ウェブ(WWW)では、ノードはインターネット上の複数のサーバマシンに分散して存在しており、各サーバに蓄積されているノードのデータはハイパーテキスト・トランスファ・プロトコル(HyperText Transfer Protocol 以下HTTPと表わす。)と呼ばれる手順に従って転送される。ノードになっているマルチメディアドキュメントは、ハイパーテキスト・マークアップ・ランゲージ(HyperText Markup Language 以下HTMLと表わす。)と呼ばれるハイパーメディア記述言語形式で記述されている。

【0007】ノード・リンク探索装置は、ネットワークを介して取得したHTMLで記述されているノードの内容を解析して、次のリンク先のノードの格納場所を表わすハイパーリンクを抽出する。ハイパーリンクは、HTMLで規定されているその開始位置を示す所定の文字列(これをタグと呼ぶ。)と終了位置を示すタグを検索することで、これらの間の文字列として抽出される。

【0008】リンク先ノードを表わすハイパーリンクは、ユニフォーム・リソース・ロケータズ(Uniform Resource Locators 以下URLと表わす。)と呼ばれる表記形式により記述されている。取得したノードのテキスト中からハイパーリンクを示すタグに挟まれた部分を見出し、この中からURLで記述された文字列を抽出することによってノードとハイパーリンクの探索を行うことができる。ノード・リンク探索装置は、このような探索を繰り返すことによって、ノード・リンク情報データベースを構築する。ノード・リンク探索装置としては、“WWW Wanderer”、“WWW Robot”、“WWW Spider”などと呼ばれるものがある。

【0009】リンク先ノードを表わすハイパーリンクは、URLにおいて以下の構成が定義されている。
 <スキーム(Scheme)>: <スキーム特有部(Scheme-Specific-Part)>

ワールド・ワイド・ウェブ(WWW)でノードのドキュメントファイルの転送プロトコルとして用いられるハイパーテキスト・トランスファ・プロトコル(HTTP)スキームの場合には、ハイパーリンクは以下のように表される。

http://<host>:<port>:<path>?<searchpart>

【0010】ここで、“<host>”はドキュメントファイルの存在するサーバマシンのホスト名を、“<port>”は通信を行う際のポート番号を表わしている。ポート番号は省略可能であり、省略した場合には標準値“80”が指定されたものとして取り扱われる。“<path>”はサーバマシン上のドキュメントファイルの存在する場所を表わしている。また、“<searchpart>”はサーバマシンにデータを渡す場合に用いられる領域である。

10 【0011】図14は、ノードとこれらノード間を接続するハイパーリンクの一例を表わしたものである。ノード1001、ノード1002、ノード1003にそれぞれ含まれるドキュメントファイルは、HTMLで記述されている。ドキュメントファイルの内容は、各種のタグによりその内容が分類され識別可能になっている。このうちハイパーリンクを示すタグは、...で表され、“URL”に対応する部分がリンク先ノードを示している。ノード1001のドキュメントファイルでは、点線で囲んだ領域1004、1005がそれぞれハイパーリンクを表わしている。

20 【0012】ノード1001は、URLで“http://hostA/tale/TOC.html”のように表記されるノードであるとする。また、ノード1002は“http://hostB:80/tale/Introduction.html”と、ノード1003は“http://hostC:8080/tale/Birth.html”と表記されるノードであるものとする。ノード1001内のハイパーリンク1004は、ノード1002を指し示している。またノード1001内のハイパーリンク1005は、ノード1003を指し示している。このようにノードのドキュメントファイル内に登録されているハイパーリンクによって、リンク先のノードが表される。

30 【0013】図15は、ノードのドキュメント同士の関係の一例を表わしたものである。一点破線で囲んだ領域は、ノードの格納されているサーバを表わしている。領域1011はホスト名が“A”のサーバマシンを、領域1012はホスト名が“B”のサーバマシンを、領域1013はホスト名が“C”のサーバマシンをそれぞれ表わしている。

40 【0014】図14に示したノード1001~1003に含まれるドキュメントファイルのうち、HTMLで規定されたタグなどを取り除いた内容1014~1016はそのノードの格納されているサーバを表わす領域1011~1013内に表わしある。また、ノード1001のリンク先ノードがノード1002とノード1003であることをハイパーリンク1017、1018により表わしている。

【0015】次に、各ノードに登録されているハイパーリンクを基にしてノード・リンク情報データベースの構築を行うノード・リンク探索装置の構成を説明する。

50 【0016】図16は、従来から使用されているノード

・リンク探索装置の構成の概要を表わしたものである。ノード・リンク探索装置1021には、インターネットなどのネットワーク1022を介して複数のサーバマシン1023が接続されている。各サーバマシン1023には、ドキュメントファイルを格納したノードが蓄積されている。ノード・リンク探索装置1021は、サーバマシン1023からノードのファイルを収集するファイル収集部1031と、サーバマシン1023から取得したノードのファイルの中からハイパーリンクを抽出するハイパーリンク抽出部1032を備えている。

【0017】また抽出したハイパーリンクを基にしてノードの格納先などの各種属性情報およびノード相互間のリンク情報を蓄積・管理するノード・リンク情報データベース1033を有する。探索範囲制御部1034は、図示しない入力端末あるいは設定用の外部ファイルから与えられる探索範囲の条件に合致する範囲にノードの探索範囲を制限する部分である。システム制御部1035は、ノード・リンク探索装置1021内の各部の動作の流れを統括的に制御する回路部分である。

【0018】ハイパーリンク抽出部1032は、ハイパーメディア記述言語で書かれているノードのマルチメディアドキュメントファイルの内容を解析するハイパーメディア記述言語解析部1036を備えている。ハイパーメディア記述言語解析部1036は、各種タグを検出することによってハイパーリンクとリンク先ノードを抽出するようになっている。探索範囲は、サーバマシン1023の識別子としてのホスト名によって指定される。探索範囲制御部1034は、ノードの格納先のサーバマシンのホスト名と指定された探索範囲としてのホスト名とを比較するホスト名比較部1037を備えている。ノードの格納位置を表わす情報の一部としてのホスト名を比較することにより、ノードが探索範囲内のものであるか否かを判定するようになっている。

【0019】ノード・リンク探索装置1021のファイル収集部1031は、探索範囲として指定されたホスト名と一致するサーバマシン1023からネットワーク1022を通じてノードのマルチメディアドキュメントファイルを読み込む。そして、ハイパーリンク抽出部1032により読み込んだファイルからハイパーリンクとリンク先ノードの情報を抽出する。探索範囲制御部1034は抽出したリンク先ノードの格納されているサーバマシンのホスト名が探索範囲内の場合には、そのリンク先ノードに対して探索を継続する。一方、リンク先ノードの格納先のサーバマシンが探索範囲外の場合は、それより先のノードへの探索は中止する。ノード・リンク情報データベース1033には、探索範囲内のノードおよびハイパーリンクについての各種属性が登録される。

【0020】このようにホスト名を探索範囲として指定するものの他に、何ら探索範囲を指定できないノード・リンク探索装置も存在する。

【0021】

【発明が解決しようとする課題】探索範囲を指定することのできないノード・リンク探索装置は、インターネットなどネットワーク全体を探索範囲とし、ネットワーク上の全てのサーバマシンに存在する全てのノードの探索を行う。このため、ノード・リンク情報データベースには、本来、探索範囲とすべきものの以外のノードに関する不必要な情報も蓄積され、その資源を浪費するばかりでなく、検索する際の作業効率も低下していた。さらに、ノードの存在するサーバマシンやネットワークをアクセスする時間が長くなり、他の利用者のネットワーク資源の利用が制限されてしまうという問題もある。

【0022】探索範囲を指定することのできる装置であっても、従来はホスト名を単位として探索範囲を制限することしかができない。これはノードの物理的な位置で探索範囲を制限していることになる。しかしながら、ハイパーテキストはノード間の意味的な関係に基づいて順次検索するものであるため、サーバマシンを単位とする探索範囲の指定では、適切な範囲指定ができない。

【0023】たとえば、図15に示した例では、サーバ“A”1011上に目次の登録されたノード1014が存在し、サーバ“B”1012上およびサーバ“C”1013上に、目次に対応する内容の文章の登録されたノード1015、1016がそれぞれ登録されている。このような場合、目次のみではなく、書かれている文書の中身までも探索範囲としたい場合には、サーバ“A”、サーバ“B”、サーバ“C”のすべてを探索範囲のホスト名として指定しなければならない。その結果、サーバ“B”、サーバ“C”上に存在する目次と何ら意味的なつながりの無い他の多数のノードまでが探索範囲となり、不要なノードの探索が行われてしまうという問題がある。

【0024】そこで本発明の目的は、意味的なつながりのない不要なノードの探索を制限することのできるノード・リンク探索装置を提供することにある。

【0025】

【課題を解決するための手段】請求項1記載の発明では、ハイパーテキストの各ノードに含まれるリンク先のノードの名称とリンク先のノードの格納位置とを表わしたリンク情報を基にして任意のノードからリンク先のノードへの探索を順次行う際の探索範囲の制限条件を探索の起点となるノードから探索先のノードまでの間に存在するノードの数である階層数の最大値として設定する探索条件設定手段と、リンク情報の示すリンク先のノードの内容をそれを蓄積しているサーバから読み出すことを探索の起点となるノードから順に繰り返し行うファイル収集手段と、このファイル収集手段によって1つのノードの内容を読み込むごとにそのノードの含むリンク情報およびこれの示すリンク先のノードと今回読み込んだノードとの対応付けを表わす情報とを記憶するノード・リ

リンク情報記憶手段と、ファイル収集手段によって1つのノードの内容を読み込むごとにそのノードに含まれるリンク情報の示すリンク先のノードの階層数を求める階層数算出手段と、この階層数算出手段によって求めた階層数が探索条件設定手段により設定した階層数の最大値よりも大きいとき今回読み込んだノード以降にリンクされているノードの内容のファイル収集手段による読み込みを中止させる探索範囲制限手段とをノード・リンク探索装置に具備させている。

【0026】すなわち請求項1記載の発明では、探索の起点となるノードから探索先のノードまでの間に存在するノードの数を階層数とし、探索を行う範囲を制限する条件を階層数の最大値により指定する。起点となるノードからリンクをたどることによって順次ノードの内容をサーバから読み出し、読み出したノードのリンク先のノードについてその名称や格納位置およびリンク元ノードとリンク先ノードとの対応関係を登録する。また、ノードをサーバから読み出すごとに読み込んだノードのリンク先のノードについての階層数を求める。この階層数が設定した最大値以上になったとき、それ以上先へリンクをたどりノードを読み出すことを中止する。

【0027】これらにより、起点としたノードからの階層数によって探索の範囲が制限される。このように起点のノードからの階層数によって探索範囲を制限しているため、探索されるノードは起点のノードと意味的なつながりの強いもののみとなる。また階層数により探索範囲を制限できるので、必要な範囲での情報のみを収集することができる。

【0028】請求項2記載の発明では、ハイパーテキストの各ノードに含まれるリンク先のノードの名称とリンク先のノードの格納位置とを表わしたリンク情報を基にして任意のノードからリンク先のノードへの探索を順次行う際の探索範囲の制限条件を探索の起点となるノードから探索先のノードまでの間に存在するリンクの数である階層数の最大値として設定する探索条件設定手段と、リンク情報の示すリンク先のノードの内容をそれを蓄積しているサーバから読み出すことを探索の起点となるノードから順に繰り返し行うファイル収集手段と、このファイル収集手段によって1つのノードの内容を読み込むごとにそのノードの含むリンク情報およびこれの示すリンク先のノードと今回読み込んだノードとの対応付けを表わす情報とを記憶するノード・リンク情報記憶手段と、ファイル収集手段によって1つのノードの内容を読み込むごとにそのノードに含まれるリンク情報の表わすリンク先のノードへのリンクの階層数を求める階層数算出手段と、この階層数算出手段によって求めた階層数が探索条件設定手段により設定した階層数の最大値よりも大きいとき今回読み込んだノード以降にリンクされているノードの内容のファイル収集手段による読み込みを中止させる探索範囲制限手段とをノード・リンク探索装置

に具備させている。

【0029】すなわち請求項2記載の発明では、探索の起点となるノードから探索先のノードまでの間に存在するリンクの数を階層数とし、探索を行う範囲を階層数の最大値により制限している。これにより起点のノードと意味的なつながりが強いノードについての情報だけを収集することができる。また階層数で指定された探索範囲内のノードについてだけその名称や格納位置ならびにノード相互間のリンク関係を収集できる。

10 【0030】請求項3記載の発明では、探索の対象となるノードがネットワークに接続された複数のサーバに分散して格納されている。

【0031】すなわち請求項3記載の発明では、探索の対象となる各ノードは、ネットワークを介して接続された複数のサーバに分散して格納されている。探索範囲を階層数によって制限しているため、ノードが複数のサーバに分散されていても、多数のサーバから必要なノードの情報だけを収集することができる。たとえば、サーバ単位でしか探索範囲を指定できない場合には、起点とするノードから必要な階層数を越えるノードのみならず、起点のノードとリンク関係のない無関係のノードまで収集される場合もある。階層数によって探索範囲を制限することにより、探索範囲のノードが複数のサーバに分散されていても、起点とするノードから意味的なつながりのある必要範囲のノードについての情報だけを収集することができる。

【0032】請求項4記載の発明では、ファイル収集手段は、今回読み込んだノードに含まれているリンク情報の示すリンク先のノードが既に読み込んだノードと同一であるか否かを判別する同一ノード判別手段と、この同一ノード判別手段により既に読み込んだノードと同一であると判別されたときそのノードの再度の読み込みを中止する多重読入中止手段とを具備している。

【0033】すなわち請求項4記載の発明では、1度読み込んだことのあるノードの再度の読み込みを防止している。これにより、ループした範囲を繰り返し探索することを回避することができる。

【0034】

【発明の実施の形態】

40 【0035】

【実施例】図1は、本発明の一実施例におけるノード・リンク探索装置の構成の概要を表わしたものである。ハイパーメディア構造上の探索手順は、探索を開始したノードを根とする木構造と見ることができる。そこで、根を始点として各ノードのある木の深さを階層数とし、探索範囲を始点からの階層数で制限するようになっている。

【0036】ノード・リンク探索装置11には、インターネットなどのネットワーク12を介してノードを蓄積した複数のサーバマシン13が接続されている。ノード

・リンク探索装置11は、サーバマシン13からノードのファイルを収集するファイル収集部21と、サーバマシン13から取得したノードのファイルの中からハイパーリンクとリンク先ノードを抽出するハイパーリンク抽出部22を備えている。

【0037】またノード・リンク探索装置11は、抽出したハイパーリンクとリンク先ノードの各種属性情報を蓄積・管理するノード・リンク情報データベース23を有する。探索範囲制御部24は、図示しない入力端末あるいは設定用のファイルから入力される探索範囲を指定するための条件に合致するようにノードの探索範囲を制限する部分である。システム制御部25は、ノード・リンク探索装置11内の各部の動作の流れを統括的に制御する回路部分である。

【0038】ハイパーリンク抽出部22は、ハイパーメディア記述言語で書かれたノードのマルチメディアドキュメントファイルの内容を解析するハイパーメディア記述言語解析部26を備えている。ハイパーメディア記述言語解析部26は、各種タグを検出することによってハイパーリンクとリンク先ノードを抽出するようになっている。また、ハイパーリンク抽出部22は、抽出したリンク先ノードの階層数を算出する階層数算出部27を備えている。階層数算出部27は、リンク元のノードの階層数に“1”を加えたものをリンク先ノードの階層数として求める。

【0039】探索範囲は、サーバマシン13の識別子としてのホスト名と、最大階層数により指定される。探索範囲制御部24は、ノードの格納先のホスト名と探索範囲として指定されたホスト名とを比較するホスト名比較部28と、そのノードの階層数と探索範囲の制限値としての最大階層数とを比較する階層数比較部29を備えている。探索範囲制御部24は、ノードの階層数、格納先のホスト名によって探索範囲を制限する。また、同一のノードの多重読み込みを防止するために、前回の検索からの経過時間により探索の範囲を制限することを行う。

【0040】ファイル収集部21は、ネットワーク12を通じてサーバマシン13からノードのマルチメディアドキュメントファイルを読み込む。ハイパーリンク抽出部22は、ハイパーメディア記述言語解析部26の解析結果を基にして、マルチメディアドキュメントファイルからハイパーリンクとリンク先ノードを抽出する。次に、階層数算出部27は、リンク元ノードに“1”を加えることにより、リンク先ノードまたはハイパーリンクの階層数を算出する。

【0041】探索範囲制御部24は、ノードあるいはハイパーリンクの階層数が指定された最大階層数以下か否か、あるいはホスト名が指定されたものと部分一致するか否かを基に探索範囲を制限する。探索範囲内のノードおよびハイパーリンクについては、それらの属性情報がノード・リンク情報データベース23に蓄積される。ノ

ード・リンク情報データベース23内でノードについての属性を登録するテーブルをノードテーブルと呼び、リンクに関する属性を登録するテーブルをリンクテーブルと呼ぶことにする。このように、階層数によって探索範囲を制限することで、不必要なノードについての属性情報の収集およびデータベースへの登録を防いでいる。

【0042】図2は、ノードテーブルの登録内容の一例を表わしたものである。ノードテーブル31には、図の左から、ノード識別子32、スキーム(Scheme)33、ホスト名34、ポート番号35、スキーム特有部(Scheme-Specific)36、階層数37、登録日時38、探索日時39、最終更新日時41が登録される。このうちノード識別子32は、図1では示していないデータベース・マネジメント・システム(DBMS)によって、ノードの属性情報をノードテーブルに登録する際に自動的に生成される識別子である。スキーム33、ホスト名34、ポート番号35およびスキーム特有部36は、ノードの格納位置を示すURLの文字列の各部分に対応するものである。階層数37は、検索開始位置からのノードの階層数を表わす項目である。

【0043】登録日時38は、ノードがデータベースに最初に登録された日時を示す。探索日時39は、このノードが探索された最新の日時を表わしている。最終更新日時41は、ノードのマルチメディアドキュメントファイルの最終更新日時を表わしている。図中、点線42で囲んだ探索日時39と、最終更新日時41の項目は、リンク元のノードとして登録されるときに更新される属性であり、点線43で囲んだ項目は、リンク元ノードのハイパーリンクにより指し示されているリンク先ノードとして登録されるときに更新される属性である。

【0044】図3は、リンクテーブルの登録内容の一例を表わしたものである。リンクテーブル51は、図の左から、リンク識別子52、リンク元ノード識別子53、リンク先ノード識別子54が登録される。これらの項目には、DBMSで自動的に割り当てられたノードの識別番号が登録される。リンク識別子52で示されるハイパーリンクの出所元のノードのノード識別子がリンク元ノード識別子53に、ハイパーリンクの指す先のノードのノード識別子がリンク先ノード識別子54として登録される。

【0045】ノード・リンク情報データベースのノードテーブル31には、探索の開始点となるノードの情報を予め少なくとも1つ登録しておく。この際、登録しておくべき項目は、ノード識別子32、スキーム33、ホスト名34、ポート番号35、スキーム特有部36、階層数37、登録日時38である。探索の起点となるノードの階層数には“0”を設定しておく。

【0046】図4は、ノード・リンク探索装置の行う動作の流れを表わしたものである。一点破線61で囲んだステップの処理は、ノード・リンク装置のシステム制御

部25によって行われる。一点破線62で囲んだステップの処理は探索範囲制御部24により、一点破線63で囲んだステップの処理はファイル収集部21およびハイパーリンク抽出部22によって行われる。まず、探索に先立ちシステム制御部は探索条件の入力などのシステムの初期化を行う(ステップS101)。探索条件は、図1で示していない入力端末や外部ファイルから入力される。探索条件として、ここでは探索範囲とすべきノードの階層数の最大値(最大階層数)と、前回の探索結果を古いものと判定するための経過時間と、探索対象とするサーバマシンを特定するための文字列(探索するホスト名の一部あるいは全部)が条件として設定される。

【0047】システム制御部25は、ノードテーブル31を検索して、未探索のノードが存在するか否かを調べる(ステップS102)。ノードテーブル31中に登録されている探索日時39が、未定義(未登録)の場合、探索日時39が現在時刻よりも探索条件として設定された経過時間以上過去の時刻である場合のいずれかに該当するとき未探索のノードと判定する。未探索のノードが存在する場合には(ステップS102; Y)、探索範囲制御部24は、探索されていない1つのノードの属性情報をノードテーブル31から取り出す。この際、階層数の小さいものから優先的にノードを選択する(ステップS103)。したがって、最初は、階層数が“0”である探索開始点のノードの属性情報が取り出される。

【0048】次に、選択したノードが指定された探索範囲内に存在するノードであるかどうかを判別する(ステップS104)。探索範囲内か否かを判別する処理については後に詳細に説明する。探索範囲内に存在するノードである場合には(ステップS104; Y)、そのノードのドキュメントファイルをネットワークを通じて読み出し、これに記述されているリンク先ノードをノードテーブルに追加登録する等のテーブル情報更新処理(ステップS105)を行う。テーブル更新処理の詳細については後に説明する。1つのノードについての更新処理(ステップS105)を終えた後、再びステップS102に戻り、未探索のノードについての探索処理を繰り返す。

【0049】選択したノードが指定された探索範囲外のノードである場合には(ステップS104; N)、ノードテーブル31内の選択したノードについての探索時刻39を現在時刻に変更し(ステップS106)、ステップS102に戻る。未探索のノードがノードテーブルに存在しなくなったとき(ステップS102; N)、探索を終了するための終了処理を行い(ステップS107)、処理を終了する(エンド)。

【0050】図5は、探索範囲内に存在するノードであるか否かを判定する際の処理の流れを表わしたものである。ここでは、階層数のみによって探索範囲内か否かを判別している。まず探索範囲制御部24は、選択したノ

ードの階層数と探索条件として設定されている最大階層数とを比較する(ステップS201)。ノードの階層数が最大階層数よりも大きい場合には(ステップS201; N)、このノードは探索範囲に存在しないものと判定する(ステップS202)。ノードの階層数が最大階層数より小さいかあるいは等しい場合には(ステップS201; Y)、このノードが探索範囲に存在するものと判定する(ステップS203)。

【0051】図6は、探索範囲内に存在するノードであるか否かを判定する処理の他の一例の流れを表わしたものである。ここでは、階層数による判別の他に、ホスト名による判別を加えている。探索範囲制御部24は、選択したノードの階層数と最大階層数とを比較し(ステップS301)、ノードの階層数が最大階層数よりも大きい場合には(ステップS301; N)、このノードは探索範囲に存在しないものと判定する(ステップS302)。ノードの階層数が最大階層数より小さいかあるいは等しい場合には(ステップS301; Y)、探索範囲のホスト名として指定された文字列と、ノードのホスト名とが部分一致するか否かを調べる(ステップS303)。

【0052】たとえば、探索文字列として“AB”が指定されたときには、“ABC”や“ABD”などホスト名の先頭から指定された文字列“AB”を含むものはすべて部分一致していると判別される。“CAB”などのように部分一致していない場合には(ステップS303; N)、このノードを探索範囲に存在しないものと判定する(ステップS302)。部分一致する場合には(ステップS303; Y)、このノードが探索範囲に存在するものと判定する(ステップS304)。ここでは、階層数の判定を行ってからホスト名の部分一致を判定したが、これらの順序を入れ換えて行ってもよい。

【0053】図7は、図4に示したテーブル更新処理の流れを表わしたものである。この処理は、ファイル収集部21とハイパーリンク抽出部22により行われる。まず、図4のステップS104により探索範囲内に存在すると判別されたノードの内容をネットワークを通じてそれが格納されているサーバマシンから読み出すことを行う。このため、ファイル収集部21は、該当するノードの属性情報をノード・リンク情報データベース23のノードテーブル31から読み込む(ステップS401)。読み込んだノードの階層数をここでは、仮にnとする。次に、ファイル収集部21は、サーバマシンとの通信に用いる図示しない通信用バッファと一時バッファを初期化した後サーバマシン13との接続を行い、ノードのドキュメントファイルの転送(読み込み)準備を行う(ステップS402)。

【0054】続いてファイル収集部21は、分散ハイパーメディアシステムの転送プロトコルに従って、ドキュメントファイルの転送を開始する。ファイル収集部21

は、転送プロトコルメッセージヘッダ部を読み込み（ステップS403）、ドキュメントファイルの記述言語などのフォーマットを調べる（ステップS404）。読み込んだファイルがハイパーメディア記述言語形式でない場合には（ステップS404；N）、そのドキュメントファイルの終端まで読み込む（ステップS405）。そして、通信用バッファおよび一時バッファの解放ならびにサーバマシンとの接続を断するなどの後処理を行い（ステップS406）、処理を終了する（エンド）。

【0055】読み込んだファイルがハイパーメディア記述言語形式の場合には（ステップS404；Y）、ファイル収集部21は、転送プロトコルメッセージ本体のドキュメントファイルを通信用バッファに読み込む（ステップS407）。次に、ハイパーメディア記述言語解析部26により、ハイパーリンクの部分を表わすタグの開始文字から終了文字までを一時バッファに移動させる（ステップS408）。一時バッファに格納したハイパーリンクの中からリンク先ノードを記述している部分を取り出す（ステップS409）。階層数算出部は、リンク元のノードの階層数“n”に“1”を加えた“n+1”を、リンク先ノードの階層数として求める（ステップS410）。

【0056】このようにして得たノードとハイパーリンクの属性情報を、ノード・リンク情報データベース23内のノードテーブル31およびリンクテーブル51に書き込む（ステップS411）。この際、リンク元ノードの属性情報、リンク先ノードの属性情報の各一部項目と、ハイパーリンクの属性情報（リンクテーブル）の全項目を書き込む。これにより、今回読み込んだノードにリンクされている1つ階層の進んだリンク先ノードについての属性情報と、読み込んだノードからそのリンク先ノードへのハイパーリンクを各テーブルに1つ追加登録したことになる。ただし、同一のノードについて既に登録されている場合には、テーブルへの追加登録は行わず、探索日時などの更新のみを行う。これにより同一のノードやハイパーリンクの多重登録が回避される。

【0057】次に、今回読み込んだノードのドキュメントファイルの終端まで処理を行ったか否かを調べ（ステップS412）、終端に到らないときは（ステップS412；N）、ステップS407に戻る。各ノードは、複数のリンク先ノードを有することがあるので、ファイルの終端までこのような処理を繰り返し行うことにより、今回読み込んだノードのリンク先ノードの全てについてノードテーブルとリンクテーブルへの登録を行う。ファイルの終端まで処理したときは（ステップS412；Y）、このノードについてのテーブル更新処理を終了する（エンド）。

【0058】ここで、図7のステップS411においてノードテーブルとリンクテーブルに登録する属性情報の内容について説明する。サーバマシンから読み込んだノ

ードをリンク元ノードとし、このノードに含まれるハイパーリンクに記述されているノードをリンク先ノードとする。まず、リンク元ノードについては、探索日時39と、最終更新日時41を登録する。これにより、ノードをいつ検索したかの最新の時刻情報を残すことができる。たとえば、図2に示すノードテーブル31において、ノード識別子が“1”のノードをサーバマシンから読み出し、このハイパーリンクの指す先としてノード識別子が“2”のノードが記述されていたものとする。この場合は、ノード識別子が“1”のノードがリンク元ノードであり、ノード識別子が“2”のノードがリンク先ノードである。

【0059】図2において点線42で囲まれている探索日時と最終更新日時がリンク元ノードの属性情報として更新される。リンク先ノードの属性情報として登録されるのは図2の点線44で囲まれている部分である。すなわち、ノード識別子と、ノードの格納位置を示すURLの文字列に対応した、スキーム、ホスト名、ポート番号、スキーム特有部、階層数算出部で求めた階層数、およびデータベースに登録された日時である登録日時である。ノード識別子は、DBMSによって自動生成されたものが登録される。リンク先のノードについては、まだ実際にサーバマシン13から読み出してそれに含まれるハイパーリンクを調べていないので、探索日時および最終更新日時は未登録のままとなる。

【0060】リンクテーブル51には、DBMSによって自動生成されたリンク識別子52と、リンク元ノード識別子53と、リンク先ノード識別子54が登録される。図3を例に説明する。ノード識別子が“1”のノードを読み込み、このハイパーリンクによってノード識別子が“2”のノードがリンク先ノードとなっているものとする。まず、リンク識別子が“1”のハイパーリンク（55）の出所元のノードのノード識別子は“1”であるので、この値をリンク元ノード識別子（56）として登録する。またハイパーリンクの指す先のノードのノード識別子は“2”であるので、この値をリンク先ノード識別子（57）として登録する。読み込んだノードの他のハイパーリンクが登録されている場合には、それらリンクについてもリンク識別子を割り当て、リンク元ノード識別子とリンク先ノード識別子が登録される。2つハイパーリンクが存在している場合には、図3の点線58で示した範囲の情報がリンクテーブル51に登録される。

【0061】ノードテーブル31およびリンクテーブル51の更新は、1つのノードをサーバマシン13から読み込み、新たなリンク先を見出すたびに行われる。したがって、図4の流れ図において1つのノードについてテーブル更新処理（ステップS105）を終えた後、ステップS102に戻ると、テーブル更新処理において新たに登録されたノードもその探索対象になる。これによ

10

20

30

40

50

り、次々とリンク先への探索が進められる。ただし、その探索範囲は階層数などの条件によって制限される。

【0062】また、探索日時が未定義か経過時間以上過去の時刻であることを未探索判定基準にしており、また一度探索したノードの探索日時には現在時刻に近い時刻が登録されるので、同一のノードについて重ねてリンク先の調査が行われることはない。これにより、たとえばノード“A”のリンク先がノード“B”で、ノード“B”のリンク先がノード“A”のようにリンクによりループが形成されている場合であっても、ノード“A”を再度調べることが無く、最大階層数までループを繰り返したようなことがない。

【0063】このようにノードの階層数によって探索の範囲を制限しているため、起点となるノードと意味的なつながりの強いノードだけを探索することができる。また、サーバ単位でしか探索範囲を制限できない場合に比べて不必要な探索を低減することができる。

【0064】変形例

【0065】これまで説明した実施例では、ノードの属性情報として階層数を持たせているが、変形例ではハイパーリンクの階層数を基にして探索範囲内か否かを判断するようになっている。装置の構成は図1に示したものと同一でありその説明を省略する。

【0066】図8は、変形例のノード・リンク探索装置で用いられるノードテーブルの登録内容の一例を表わしたものである。図2と同一の項目には同一の符号を付してあり、それらの説明を適宜省略する。ノードテーブル71は、図2に示したノードテーブル31に比べて、階層数と探索日時を登録する項目が削除されている点で相違する。

【0067】図9は、変形例のノード・リンク探索装置で用いるリンクテーブルの登録内容の一例を表わしたものである。図3と同一項目には同一の符号を付してあり、それらの説明を適宜省略する。リンクテーブル81は、図3に示したリンクテーブル51に加えて、階層数の項目82と、探索日時の項目83を備えている。これらは、実施例においてはノードテーブルに登録されていたものである。変形例では、探索に先立って、探索の開始点となるハイパーリンクを少なくとも1つリンクテーブル81に登録しておく必要がある。ここで登録される探索開始点となるハイパーリンクは、仮想的なものであり、リンク元ノードが存在しない（不明）。また、探索の開始点となるハイパーリンクの示すリンク先ノードについての属性情報をノードテーブル71に予め登録しておかなければならない。

【0068】探索に先立ってノードテーブル71に登録する内容として、まず探索の起点のハイパーリンクの示すリンク先ノードのノード識別子がある。この値はDBMSによって自動的に割り当てられる。さらに、ノードの格納位置を示すURLの文字列に対応した項目とし

て、スキーム、ホスト名、ポート番号、スキーム特有部を初期登録しておく。また、このノードがデータベースに登録された時刻を表わす登録日時を初期登録する。図8の例では、点線72で囲んだ範囲の項目が初期登録される。

【0069】リンクテーブル81に初期登録しておく内容としては、探索の起点となるハイパーリンクの識別子であるリンク識別子がある。この値は、DBMSによって自動生成される。探索の起点となるハイパーリンクのリンク元は不明であるため、リンク元ノード識別子の初期値は未定義を表わす“0”とする。リンク先ノード識別子は、リンク先となるノードに対してDBMSの割り当てたノード識別子と同一の値を登録しておく。階層数は、探索の開始点である“0”を初期登録する。また探索日時は、当該ハイパーリンクについての探索を行った日時を登録するものであり、探索開始の初期値としては未定義のままとする。図9の例では点線84で囲んだ項目が初期登録される。

【0070】図10は、変形例におけるノード・リンク探索装置の行う処理の流れを表わしたものである。一点破線101で囲まれたステップは、システム制御部25の行う処理を表わしている。一点破線102で囲まれたステップは、探索範囲制御部24により、一点破線103で囲まれたステップは、ファイル収集部21およびハイパーリンク抽出部22によって行われる処理を表わしている。

【0071】まず、システム制御部25は、当該システムの初期化を行う（ステップS501）。この際、探索範囲とするハイパーリンクの階層数の最大値としての最大階層数と、前回の探索結果を古いものとして扱う基準となる経過時間とを設定する。さらにホスト名によってサーバの範囲を設定する場合には、ホスト名を制限するための文字列を入力する。これらは、図示しない入力端末あるいは外部ファイルから取り込む。次に、リンクテーブル81の中に探索していないハイパーリンクが存在するか否かを調べる（ステップS502）。探索されていないハイパーリンクとは、探索日時が未定義（未登録）のもの、あるいは探索日時が現在時刻よりも経過時間以上古いものである。

【0072】探索されていないハイパーリンクがリンクテーブル81に存在するときは（ステップS502；Y）、探索されていないハイパーリンクのうちの1つを選択しその属性情報を取り出す（ステップS503）。次に、探索範囲制御部24は、選択した1つのハイパーリンクが探索範囲内であるかどうかを判定する（ステップS504）。階層数あるいは階層数とホスト名の双方により判定されるが、その詳細については後に説明する。

【0073】探索範囲内に存在する場合には（ステップS504；Y）、テーブル更新処理（ステップS50

5)を行う。この処理では、ハイパーリンクの示すリンク先ノードの属性情報をノードテーブル71から取り出し、このノードのドキュメントファイルをサーバマシンから読み出し、これに登録されたリンク先ノードの属性情報を追加登録する等を行う。処理の詳細な流れについては後述する。ハイパーリンクの示すリンク先ノードについてのテーブル更新処理(ステップS505)を終えた後、再びステップS502に戻り、未探索のハイパーリンクについての探索処理を繰り返す。

【0074】選択したハイパーリンクが探索範囲外の場合には(ステップS504;N)、リンクテーブル81中の選択したハイパーリンクについての探索日時39を現在時刻に変更し(ステップS506)、ステップS502に戻る。未探索のハイパーリンクがリンクテーブル81に存在しなくなったとき(ステップS502;N)、探索を終了するための終了処理を行い(ステップS507)、処理を終了する(エンド)。

【0075】図11は、ハイパーリンクが探索範囲内か否かを判定する際の処理の流れを表わしたものである。ここでは、ハイパーリンクの階層数のみによって探索範囲内か否かを判別している。まず、探索範囲制御部24は、選択したハイパーリンクの階層数と探索条件として設定されている最大階層数とを比較する(ステップS601)。ハイパーリンクの階層数が最大階層数よりも大きい場合には(ステップS601;N)、このハイパーリンクは探索範囲に存在しないものと判定する(ステップS602)。ハイパーリンクの階層数が最大階層数より小さいかあるいは等しい場合には(ステップS601;Y)、このハイパーリンクが探索範囲に存在するものと判定する(ステップS603)。

【0076】図12は、ハイパーリンクが探索範囲内か否かを判定する処理の他の一例の流れを表わしたものである。ここでは、ハイパーリンクの階層数による判別の他に、ホスト名による判別を加えている。探索範囲制御部24は、選択したハイパーリンクの階層数と最大階層数とを比較し(ステップS701)、ハイパーリンクの階層数が最大階層数よりも大きい場合には(ステップS701;N)、このハイパーリンクは探索範囲に存在しないものと判定する(ステップS702)。ハイパーリンクの階層数が最大階層数より小さいかあるいは等しい場合には(ステップS701;Y)、このハイパーリンクの示すリンク先ノードについての属性情報をノードテーブル71から取り出す(ステップS703)。

【0077】取り出したリンク先ノードの属性情報に含まれるホスト名が探索範囲のホスト名として指定された文字列と、部分一致するか否かを調べる(ステップS704)。部分一致する場合には(ステップS704;Y)、先のハイパーリンクが探索範囲内に存在するものと判定する(ステップS705)。部分一致しない場合には(ステップS705;N)、ハイパーリンクが探索

範囲内に存在しないと判定する(ステップS702)。ここでは、階層数の判定を行ってからホスト名の部分一致を判定したが、これらの順序を入れ換えて行ってもよい。

【0078】図13は、図10に示したテーブル更新処理の流れを表わしたものである。この処理は、ファイル収集部21とハイパーリンク抽出部22により行われる。まず、図10のステップS504により探索範囲内に存在すると判別されたハイパーリンクの属性情報をリンクテーブル81から読み込む(ステップS801)。読み込んだハイパーリンクの階層数をここでは、仮にnとする。次に、このハイパーリンクのリンク先ノード識別子に対応するノードの属性情報をノードテーブル71から読み込む(ステップS802)。

【0079】ファイル収集部21は、サーバマシンとの通信に用いる図示しない通信用バッファと一時バッファを初期化した後サーバマシンとの接続を行い、ハイパーリンクの示すリンク先ノードのドキュメントファイルの読み込み準備を行う(ステップS803)。続いてファイル収集部21は、分散ハイパーメディアシステムの転送プロトコルに従って、ドキュメントファイルの転送を開始する。

【0080】ファイル収集部21は、転送プロトコルメッセージヘッダ部を読み込み(ステップS804)、ドキュメントファイルの記述冒語がハイパーメディア記述冒語形式でない場合には(ステップS805;N)、そのドキュメントファイルの終端まで読み込む(ステップS806)。そして、通信用バッファおよび一時バッファの解放ならびにサーバマシンとの接続を断するなどの後処理を行い(ステップS807)、処理を終了する(エンド)。

【0081】読み込んだファイルがハイパーメディア記述冒語形式の場合には(ステップS805;Y)、ファイル収集部21は、転送プロトコルメッセージ本体のドキュメントファイルを通信用バッファに読み込む(ステップS808)。次に、ハイパーリンクの部分を表わすタグの開始文字から終了文字までを一時バッファに移動させ(ステップS809)、一時バッファに格納したハイパーリンクの中からリンク先ノードを記述している部分を取り出す(ステップS810)。階層数算出部27は、抽出したハイパーリンクの階層数として“n+1”を設定する(ステップS812)。

【0082】このようにして得たノードとハイパーリンクの属性情報を、ノード・リンク情報データベース23内のノードテーブル71およびリンクテーブル81に書き込む(ステップS812)。この際、リンク元ノードの属性情報、リンク先ノードの属性情報の各一部項目と、ハイパーリンクの属性情報の全項目を書き込む。これにより、今回の読み込んだノードのリンク先ノードについての属性情報と、読み込んだノードからそのリンク

先ノードへのハイパーリンクについて登録が行われる。ただし、同一のリンク元とリンク先を有するハイパーリンクが既に登録されている場合には、テーブルへの追加登録は行わず、探索日時などの更新のみを行う。これにより同一のノードやハイパーリンクが多重登録されることが回避される。

【0083】次に、今回読み込んだノードのドキュメントファイルの終端まで処理を行ったか否かを調べ（ステップS813）、終端に到らないときは（ステップS813; N）、ステップS808に戻る。各ノードは、複数のリンク先ノードを有することがあるので、ファイルの終端までこのような処理を繰り返し行うことにより、今回読み込んだノードのリンク先ノードの全てについてノードテーブルとリンクテーブルへの登録が行われる。ファイルの終端まで処理したとき（ステップS813; Y）、処理を終了する（エンド）。

【0084】ここで、図13のステップS812においてノードテーブル71とリンクテーブル81に登録される属性の内容について説明する。サーバマシンから読み込んだノードをリンク元ノードとし、このノードに含まれるハイパーリンクに記述されているノードをリンク先ノードとする。まず、リンク元ノードについては、最終更新日時41を登録する。たとえば、図8に示すノードテーブル71において、ノード識別子が“1”のノードをサーバマシンから読み出し、このハイパーリンクの指す先としてノード識別子が“2”のノードが記述されていたものとする。この場合は、ノード識別子が“1”のノードがリンク元ノードであり、ノード識別子が“2”のノードがリンク先ノードになる。

【0085】図8において点線73で囲まれている最終更新日時がリンク元ノードの属性情報として更新される。リンク先ノードの属性情報として登録されるのは図8の点線74で囲まれている部分である。すなわち、ノード識別子と、ノードの位置を示すURLの文字列に対応した、スキーム、ホスト名、ポート番号、スキーム特有部、およびデータベースに登録された日時である登録日時である。ノード識別子は、DBMSによって自動生成されたものが登録される。

【0086】リンクテーブル81には、DBMSによって自動生成されたリンク識別子と、リンク元ノード識別子と、リンク先ノード識別子と、階層数と、探索日時が登録される。図9を例に説明する。リンク識別子が“1”のハイパーリンクを基にしてリンク先であるノード識別子が“1”のノードを読み込んだものとする。読み込んだノードに記述されているハイパーリンクによりノード識別子が“2”とノード識別子が“3”のノードとがリンクされているものとする。

【0087】この際、ハイパーリンク84を参照して探索を行ったので、その時刻を探索日時85として登録する。次に新たに得られたハイパーリンクについてのリン

ク識別子をDBMSから取得し、これを追加登録するリンク識別子の欄に登録する。たとえば、リンク識別子が“2”のハイパーリンク(86)のリンク識別子(87)として“2”を登録する。この値は、テーブルの登録順などを基準にDBMSにより適宜与えられるID番号である。

【0088】ハイパーリンク86のリンク元ノードは、そのノード識別子の値が“1”であるので、リンク元ノード識別子88として“1”を登録する。またリンク先ノード識別子89には、リンク先のノードの属性をノードテーブル71に登録する際にDBMSにより与えられたノード識別子の値、すなわち“2”を登録する。また、階層数91には、階層数“0”のハイパーリンクをたどって得たノードから取得したハイパーリンクであるので“0”に“1”を加えた値“1”を登録する。ノード識別子が“3”のノードに向けてのハイパーリンク92についても同様の手順により登録される。追加登録したハイパーリンクについて図9の点線93で示した範囲の情報が登録されることになる。

【0089】ノードテーブル71およびリンクテーブル81の更新は、1つのハイパーリンクを基にしてそのリンク先ノードをサーバマシンから読み込み、新たなリンク先を見い出すたびに行われる。したがって、図10の流れ図において1つのノードについてテーブル更新処理（ステップS505）を終えた後、ステップS502に戻ると、テーブル更新処理において新たに登録されたハイパーリンクもその探索対象になる。これにより、次々とリンク先への探索が進められる。ただし、その探索範囲は階層数などの条件によって制限される。

【0090】また、探索日時が未定義か経過時間以上過去の時刻であることを未探索の判定基準にしているのので、一度探索したハイパーリンクについて再度調べることがない。これにより、たとえばノード“A”のリンク先がノード“B”で、ノード“B”のリンク先がノード“A”のようにリンクによりループが形成されている場合であっても、ノード“A”からノード“B”へのハイパーリンクを再度調べることが無く、最大階層数までループをたどるようなことがない。

【0091】このようにハイパーリンクの階層数を基に探索範囲を制限しても、ノードの階層数を基に探索範囲を制限したときと同様の効果を得ることができる。ただし、ハイパーリンクに階層数を付与する場合には、リンクテーブルを参照して得たハイパーリンクのリンク先ノードを調べるために、ノードテーブルも参照する必要がある。実施例のようにノードに階層数を付与した場合に比べてテーブルの参照処理が増加する。

【0092】

【発明の効果】以上詳細に説明したように請求項1ないし請求項3記載の発明によれば、起点のノードからの階層数によって探索範囲を制限しているので、探索される

ノードは起点のノードと意味的なつながりの強いもののみとなる。また階層数により探索範囲を制限できるので、必要な範囲での情報のみを収集することができる。また、不要なノードの探索が行われないので、サーバおよびサーバとの間の通信回線の負担を軽減できるとともに探索に要する時間を短くすることができる。

【0093】また請求項4記載の発明によれば、1度読み込んだことのあるノードを再度読み込むことを防止したので、ループした範囲を繰り返し探索することを回避することができる。これにより探索を効率良く行うことができる。

【図面の簡単な説明】

【図1】本発明の一実施例におけるノード・リンク探索装置の構成の概要を表わしたブロック図である。

【図2】ノードテーブルの登録内容の一例を表わした説明図である。

【図3】リンクテーブルの登録内容の一例を表わした説明図である。

【図4】ノード・リンク探索装置の行う動作の流れを表わした流れ図である。

【図5】探索範囲内に存在するノードであるか否かを判定する際の処理の流れを表わした流れ図である。

【図6】探索範囲内に存在するノードであるか否かを判定する処理の他の一例の流れを表わした流れ図である。

【図7】図4に示したテーブル更新処理の流れを表わした流れ図である。

【図8】変形例のノード・リンク探索装置で用いられるノードテーブルの登録内容の一例を表わした説明図である。

【図9】変形例のノード・リンク探索装置で用いられるリンクテーブルの登録内容の一例を表わした説明図である。

る。

【図10】変形例におけるノード・リンク探索装置の行う動作の流れを表わした流れ図である。

【図11】ハイパーリンクが探索範囲内に存在するか否かを判定する際の処理の流れを表わした流れ図である。

【図12】ハイパーリンクが探索範囲内に存在するか否かを判定する処理の他の一例の流れを表わした流れ図である。

【図13】図10に示したテーブル更新処理の流れを表わした流れ図である。

【図14】ノードとこれらノード間を接続するハイパーリンクの一例を表わした説明図である。

【図15】ノードのドキュメント同士の関係の一例を表わした説明図である。

【図16】従来から使用されているノード・リンク探索装置の構成の概要を表わしたブロック図である。

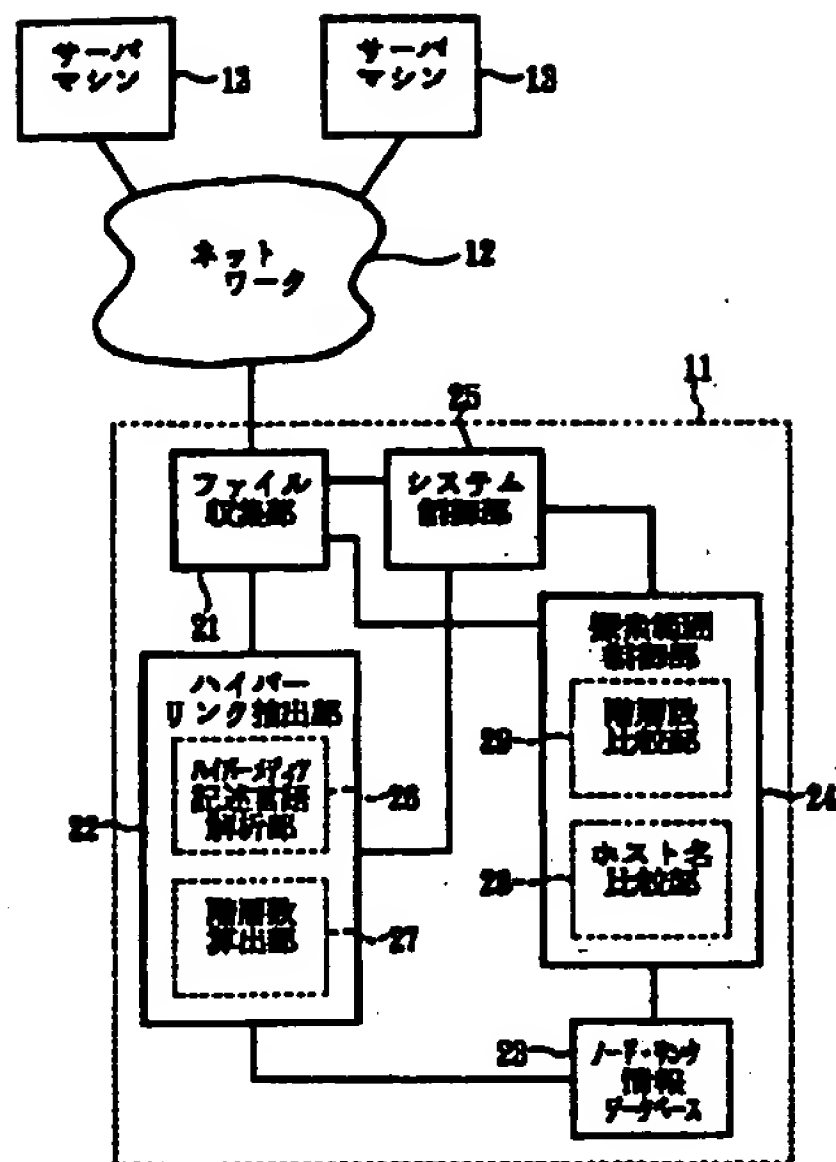
【符号の説明】

- 11 ノード・リンク探索装置
- 12 ネットワーク
- 13 サーバマシン
- 21 ファイル収集部
- 22 ハイパーリンク抽出部
- 23 ノード・リンク情報データベース
- 24 探索範囲制御部
- 25 システム制御部
- 26 ハイパーメディア記述言語解析部
- 27 階層数算出部
- 28 ホスト名比較部
- 29 階層数比較部
- 31、71 ノードテーブル
- 51、81 リンクテーブル

【図2】

ノード識別子	24-1	24-2	24-3	24-4	24-5	24-6	24-7	24-8
ノード識別子	URL	ホスト名	ポート番号	スキーム特有部	階層数	登録日時	探索日時	最終更新日時
1	http	host A	80	TOC.html	0	19951101224731	19951103193947	19951028135209
2	http	host B	80	Introduction.html	1	19951103193948		43
m	http	host X	10080	Special.html	n			

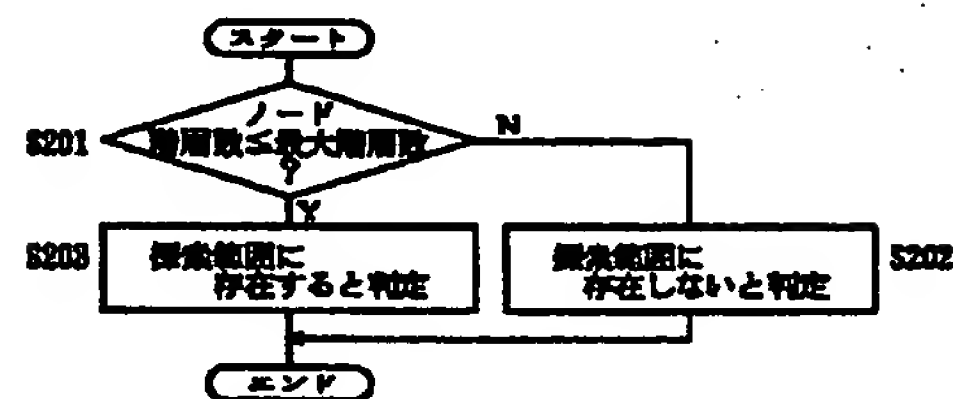
【図1】



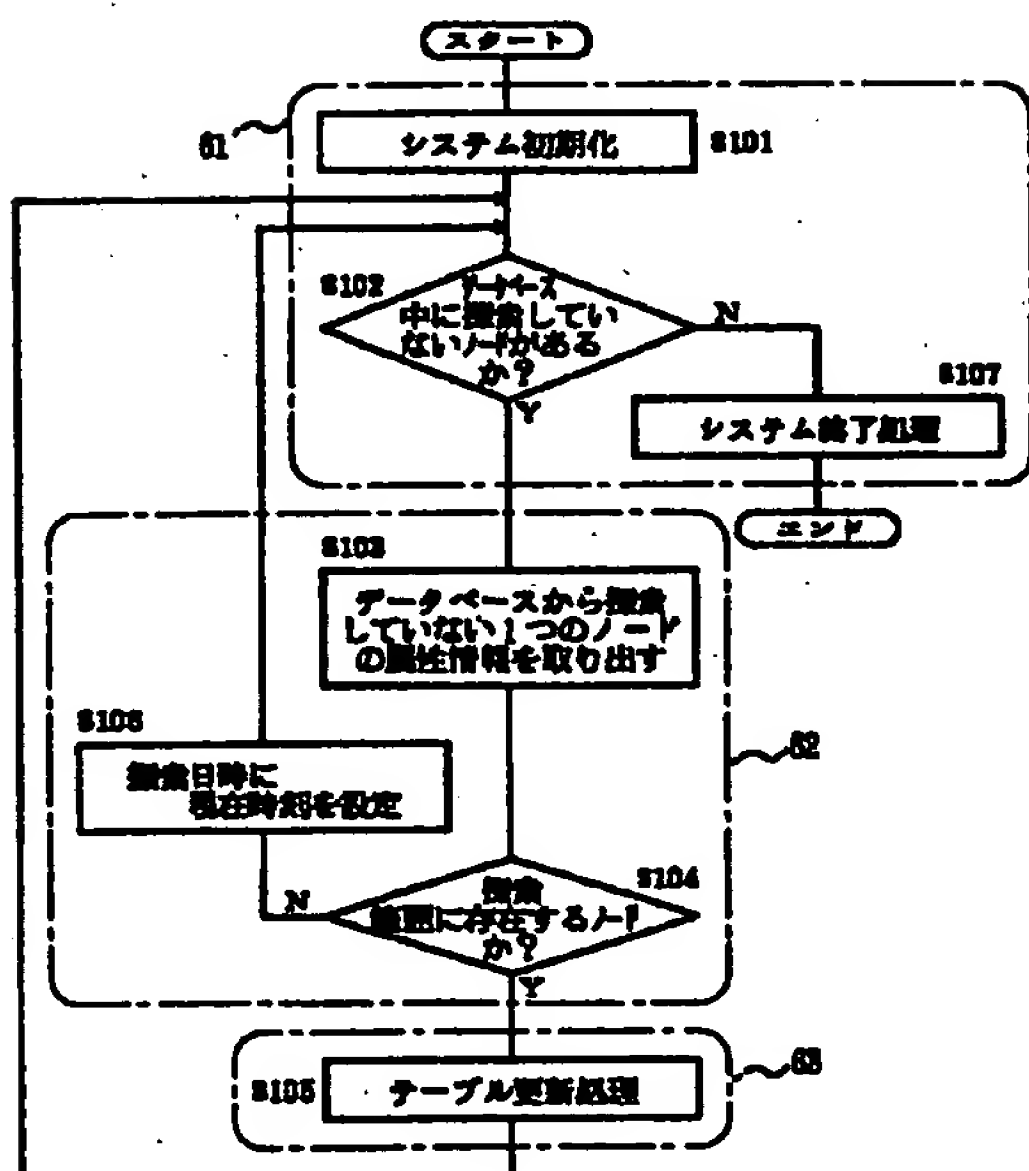
【図3】

リンク 識別子	リンク元 ノード識別子	リンク先 ノード識別子
1	1	2
2	1	3
	58	
m	j	k

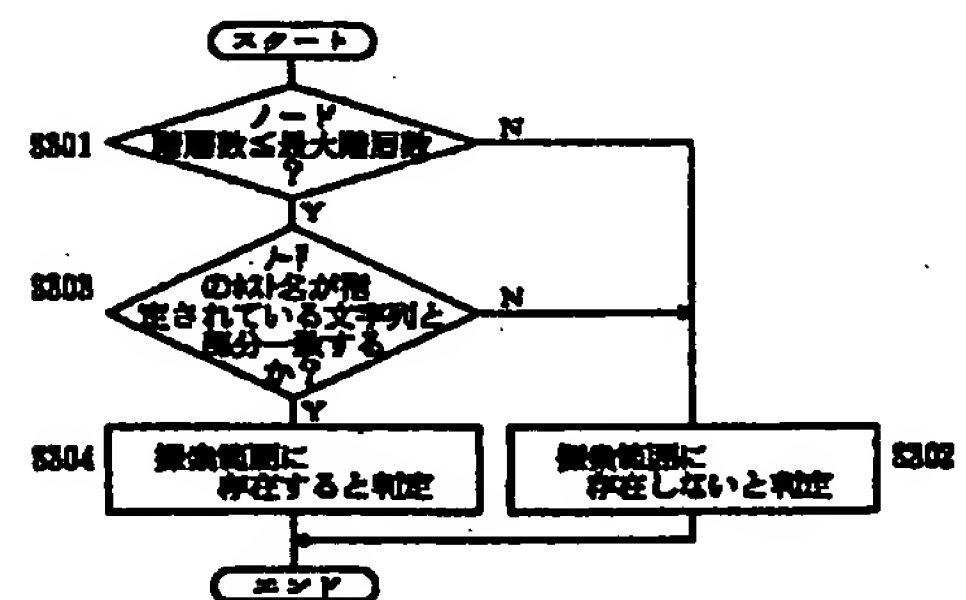
【図5】



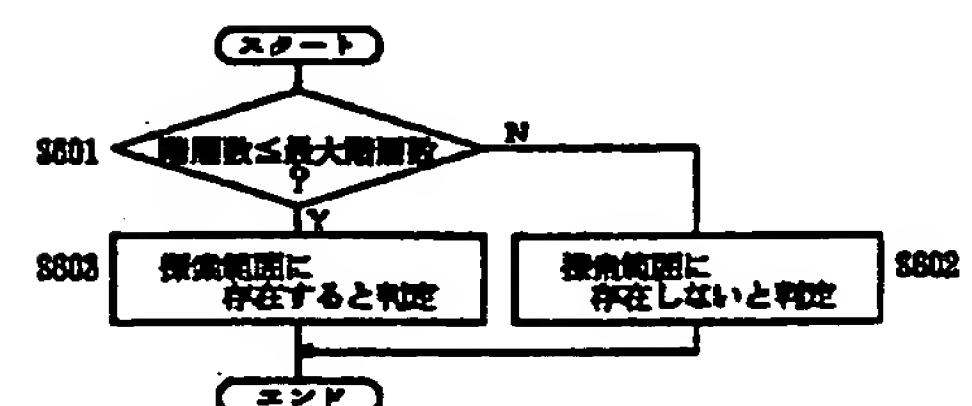
【図4】



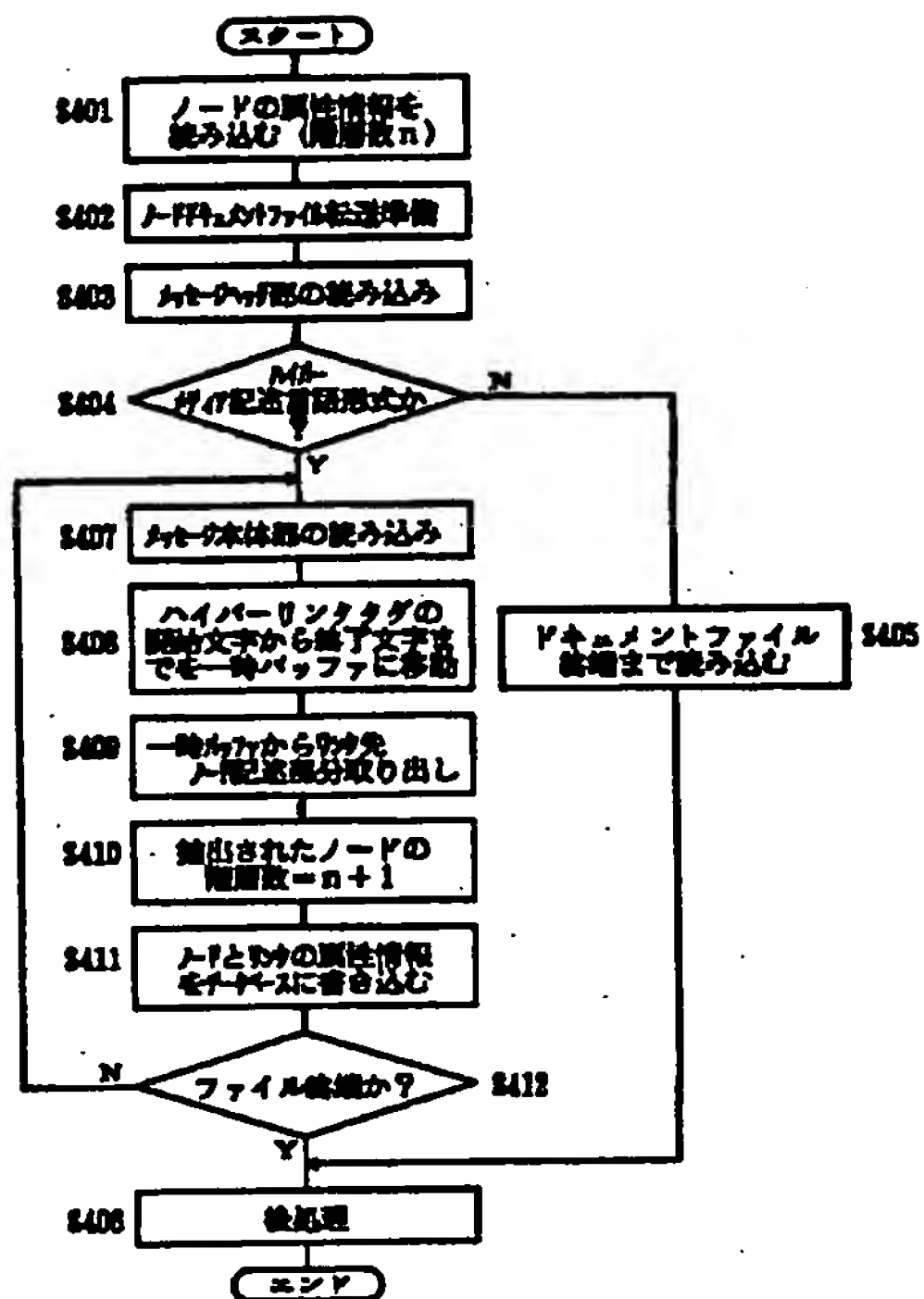
【図6】



【図11】



【図7】



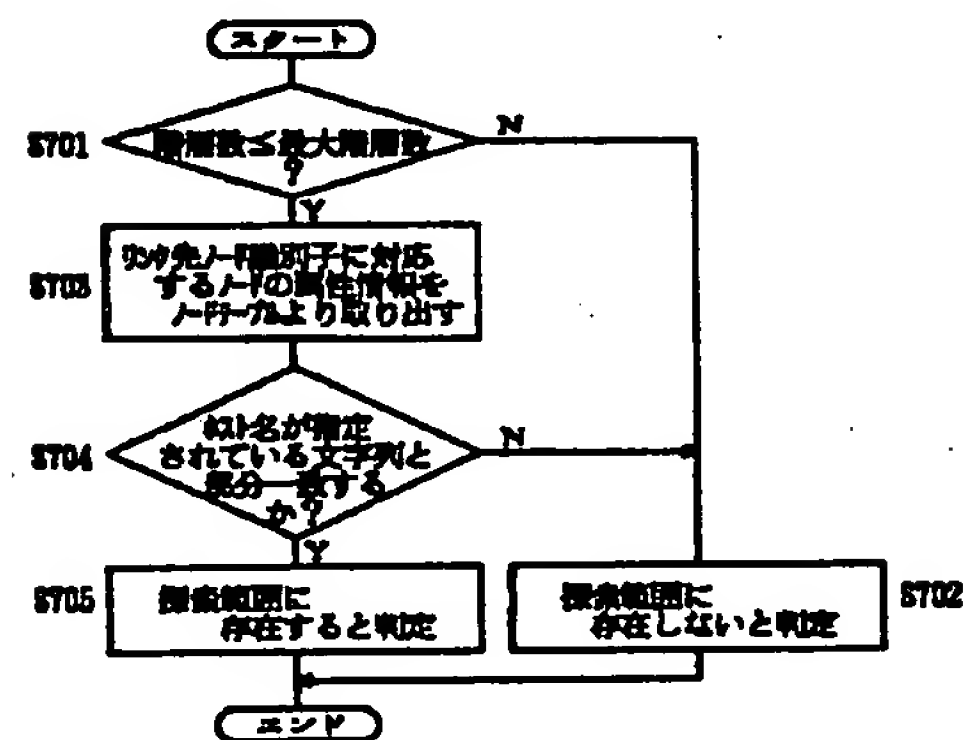
【図8】

32	33	34	35	36	37	41
ノード識別子	URL	URL名	ポート番号	スキーム特有部	登録日時	最終更新日時
1	http	hostA	80	TOC.html	19951101234731	19951028185209
2	http	hostB	80	Introduction.html	19951108183948	79
m	http	hostK	10080	Special.html		

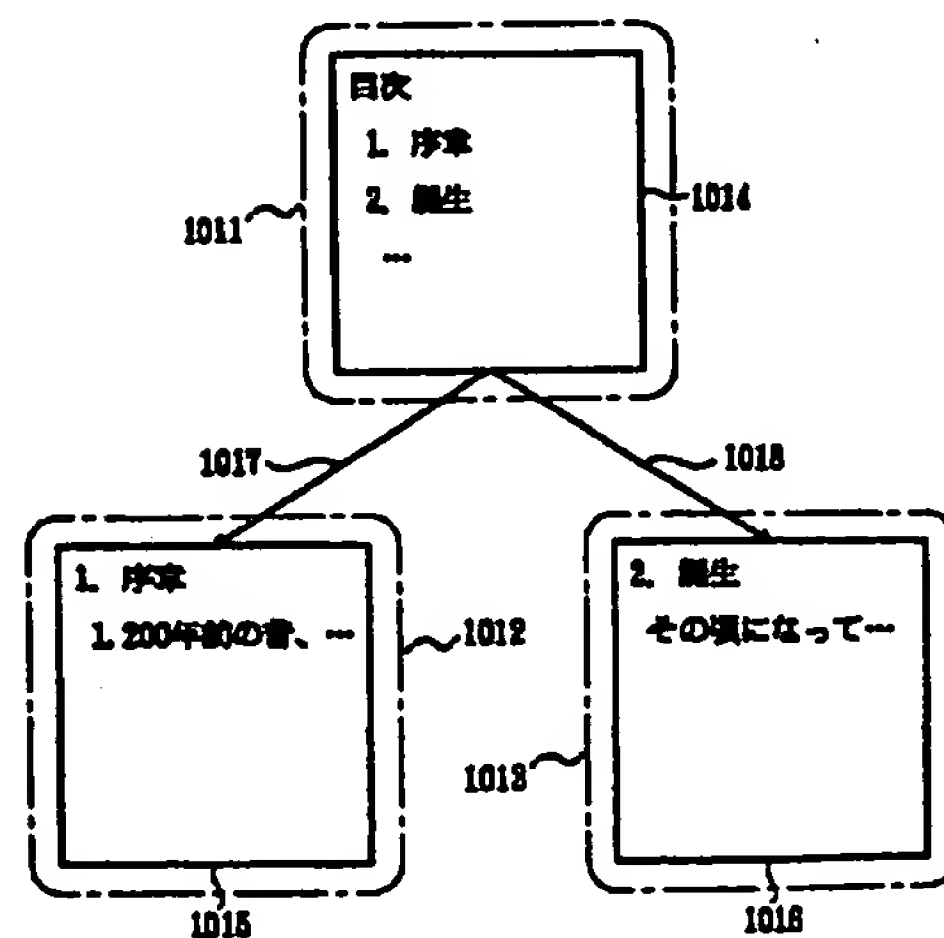
【図9】

32	33	34	37	38
リンク先ノード識別子	リンク先ノード識別子	リンク先ノード識別子	階層数	登録日時
1	0	1	0	19951102193947
2	1	2	1	81
3	1	3	1	85
87	88	89		
m	j	k	23	

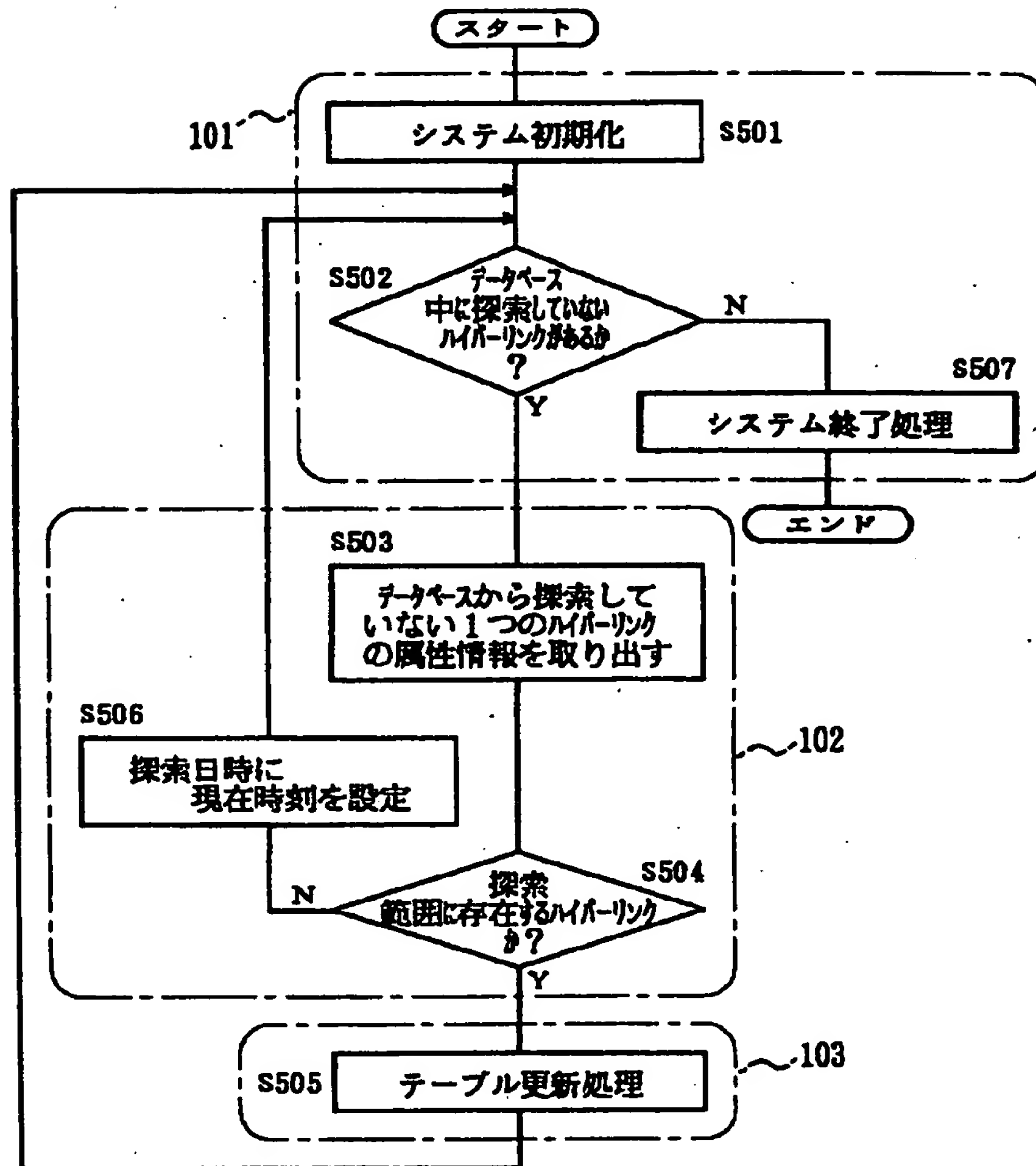
【図12】



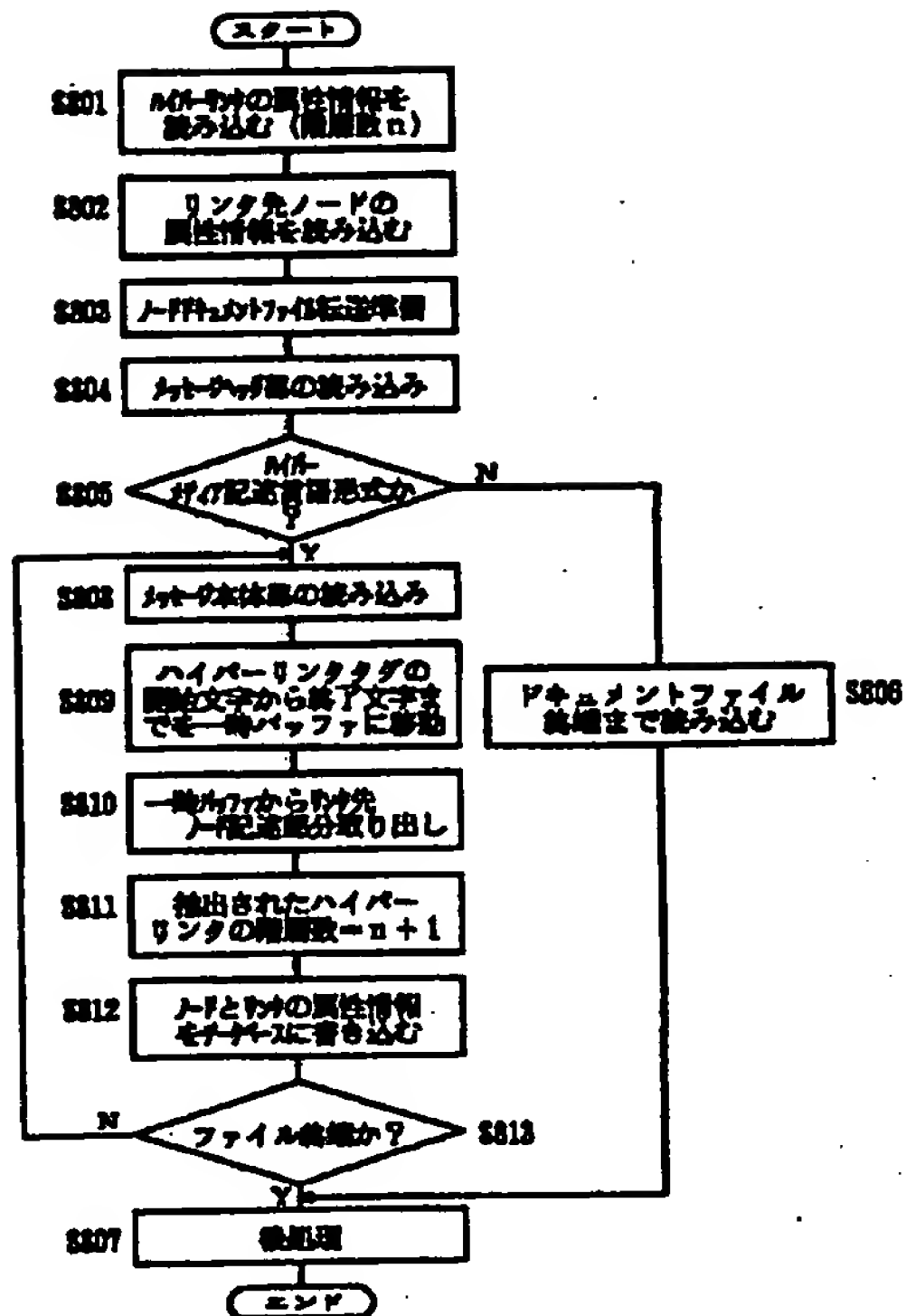
【図15】



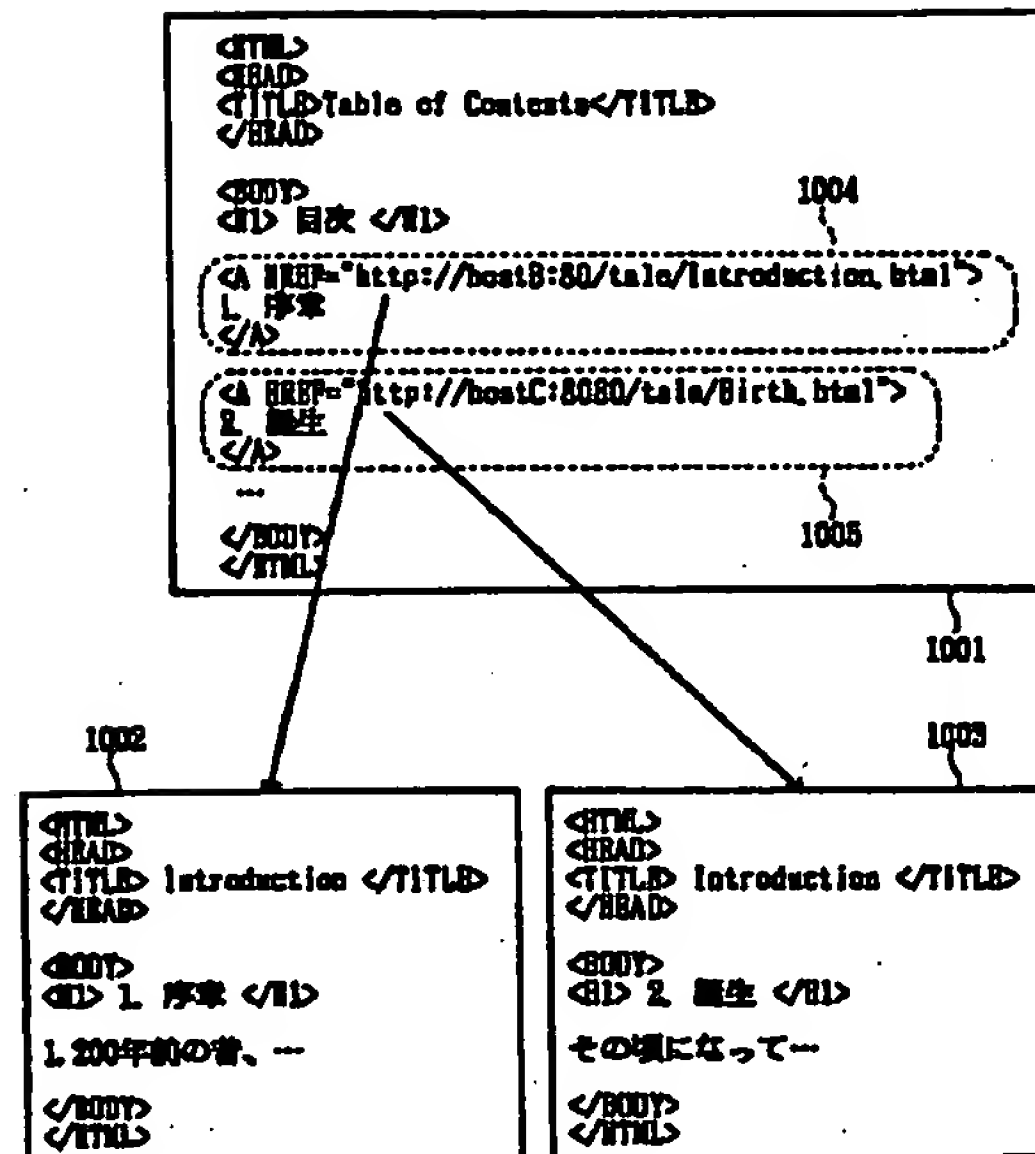
【図10】



【図13】



【図14】



【図16】

